Data Dictionary for Water Quality Statistical Summary

National Stormwater BMP Database

Prepared by

GeoSyntec Consultants

June 2003

ASCE

American Society of Civil Engineers

EPA

United States
Environmental Protection
Agency

Disclaimer

The National Stormwater BMP Database is intended to provide a consistent and scientifically defensible set of data on BMP designs and related performance. Although the database team has made an extensive effort to assess the quality of the data entered for consistency and accuracy, the use of the database information or any analysis results provided by the project team is solely at the risk and option of the user.

The database team's tasks have not included, and will not include in the future, recommendations of one BMP type over another. However, the teams' tasks have included reporting on the performance characteristics of BMPs based upon the entered data and information in the database, including peer reviewed performance assessment techniques. How the public and private sector use this information is beyond the Database team's influence or control. The intended purpose of the database is to provide a data exchange tool that permits characterization of BMPs solely upon their measured performance using the same protocols for measurements and reporting information.

The Database team does not endorse any BMP over another and any assessments of performance by others should not be interpreted or reported as the recommendations of the Database team.

Overview

This document is intended to be used in conjunction with the Excel Spreadsheet "NSW Statistical Summary 6-18-03.xls" as a reference. The table included below provides a data dictionary for the column headers contained in the referenced file. The information included in the file is a result of work conducted in statistically analyzing data stored in the National Stormwater Best Management Practices Database.

Table 1. Data Dictionary for Water Quality Statistical Summary Table (Updated 6-18-03)

Record Identifiers	
State	State where test was performed
City	City closest to the test site. The site does not have to be within the city limits.
BMP Category	BMP Type Category
BMP Description	Description of BMP Type Category
BMP ID	Unique ID number assigned to each individual BMP (Link to NSW database)
BMP name	Name of the BMP
Parameter name	Common name for water quality constituent
Parameter ID	Unique ID number assigned to each individual WQ constituent
Summary Statistics for Inflow Data	
Inflow n	Number of inflow EMCs – no statistics calculated for samples n<3
Raw Inflow Mean	Mean of inflow EMCs
Raw Inflow StDev	Standard deviation of inflow EMCs
Raw Inflow CV	Coefficient of variation of inflow EMCs
LN Inflow Mean	Mean of the natural log of inflow EMCs
LN Inflow StDev	Standard deviation of the natural log of inflow EMCs
LN Inflow CV	Coefficient of variation of the natural log of inflow EMCs
Est Inflow Mean	Arithmetic estimate of the mean inflow EMCs (for log-normally distributed data this estimate is a good estimate of the central tendency of the data)
Est Inflow StDev	Arithmetic estimate of the standard deviation of inflow EMCs (for log-normally distributed data this estimate is a good estimate of the dispersion in the data)
Inflow LCL	Land's lower 95% confidence limit of the inflow mean (this value should be used in conjunction with the EST Inflow Mean)
Inflow UCL	Land's upper 95% confidence limit of the inflow mean (this value should be used in conjunction with the EST Inflow Mean)
Shapiro-Wilks Test for Normality and Log-normality of Inflow	
Raw Inflow diff (SW)	Difference between Shapiro-Wilks W-statistic and the corresponding test quantile for inflow untransformed data.
Indicate Inflow N (SW)	Interpretation of the Shapiro-Wilks test for untransformed inflow data – "Yes" if the null hypothesis (data is normally distributed) cannot be rejected, "No" if the null hypothesis can be rejected.
LN Inflow diff (SW)	Difference between Shapiro-Wilks W-statistic and the corresponding test quantile for log-transformed inflow data.
Indicate Inflow LN (SW)	Interpretation of the Shapiro-Wilks test for LN transformed inflow data – "Yes" if the null hypothesis (data is log-normally distributed) cannot be rejected, "No" if the null hypothesis can be rejected.
Lilliefors Test for Normality and Log-normality of Inflow	
Raw Inflow Lill Prob (KS)	P-value (SL=0.05) for the Lilliefors' form of the Kolmogorov-Smirnov test of normality for untransformed inflow data.
Indicate Inflow N (KS)	Interpretation of the KS test for untransformed inflow data – "Yes" if the null hypothesis (data is normally distributed) cannot be rejected, "No" if the null hypothesis can be rejected.
LN Inflow Lill Prob (KS)	P-value (SL=0.05) for the Lilliefors' form of the Kolmogorov-Smirnov test of normality for log-transformed data.
Indicate Inflow LN (KS)	Interpretation of the KS test for log-transformed data – "Yes" if the null hypothesis (data is log-normally distributed) cannot be rejected, "No" if the null hypothesis can be rejected.

Table 1 cont. Data Dictionary for Water Quality Statistical Summary Table (Updated 6-18-03)

Summary Statistics for Outflow Data	
Outflow n	Number of outflow EMCs – no statistics calculated for n<3
Raw Outflow Mean	Mean of outflow EMCs
Raw Outflow StDev	Standard deviation of outflow EMCs
Raw Outflow CV	Coefficient of variation of outflow EMCs
LN Outflow Mean	Mean of the natural log of outflow EMCs
LN Outflow StDev	Standard deviation of the natural log of outflow EMCs
LN Outflow CV	Coefficient of variation of the natural log of outflow EMCs
Est Outflow Mean	Arithmetic estimate of the mean outflow EMCs (for log-normally distributed data this estimate is a good estimate of the central tendency of the data)
Est Outflow StDev	Arithmetic estimate of the standard deviation of outflow EMCs (for log-normally distributed data this estimate is a good estimate of the dispersion in the data)
Outflow LCL	Land's lower confidence limit of the outflow mean (this value should be used in conjunction with the EST Outflow Mean)
Outflow UCL	Land's upper confidence limit of the outflow mean (this value should be used in conjunction with the EST Outflow Mean)
Shapiro-Wilks (SW) Test for Normality and Log-normality of Outflow	
SW SL	Significance level used for Shapiro-Wilks test for normality.
Raw Outflow diff (SW)	Difference between Shapiro-Wilks W-statistic and the corresponding test quantile for untransformed outflow data.
Indicate Outflow N (SW)	Interpretation of the Shapiro-Wilks test for untransformed outflow data – "Yes" if the null hypothesis (data is normally distributed) cannot be rejected, "No" if the null hypothesis can be rejected.
LN Outflow diff (SW)	Difference between Shapiro-Wilks W-statistic and the corresponding test quantile for log-transformed outflow data.
Indicate Outflow LN (SW)	Interpretation of the SW test for log-transformed data – "Yes" if the null hypothesis holds, "No" if the null hypothesis is rejected.
Kolmogorov-Smirnov (KS) Lilliefors Test for Normality and Log-normality of Outflow	
Raw Outflow Lill Prob (KS)	P-value (SL=0.05) for the Lilliefors' form of the Kolmogorov-Smirnov test of normality for untransformed outflow data.
Indicate Outflow N (KS)	Interpretation of the KS test for untransformed outflow data – "Yes" if the null hypothesis (data is normally distributed) cannot be rejected, "No" if the null hypothesis can be rejected.
LN Outflow Lill Prob (KS)	P-value (SL=0.05) for the Lilliefors' form of the Kolmogorov-Smirnov test of normality for log-transformed outflow data.
Indicate Outflow LN (KS)	Interpretation of the Shapiro-Wilks test for LN transformed outflow data – "Yes" if the null hypothesis (data is log-normally distributed) cannot be rejected, "No" if the null hypothesis can be rejected.

Table 1 cont. Data Dictionary for Water Quality Statistical Summary Table (Updated 6-18-03)

Kruskal-Wallis (KW) Test (Wilcoxon Rank Sum test) – Nonparametric test for shift in central tendency between inflow and outflow	
KW SL	Significance level used for the Kruskal-Wallis rank sum test of variance.
Raw p (KW)	Reported p-value for the KW test for untransformed data.
Raw Sig Diff (KW)	Interpretation of the KS test for untransformed data – "Yes" if the null hypothesis holds, "No" if the null hypothesis is rejected.
LN p (KW)	Reported p-value for the KW test for log-transformed data.
LN Sif Diff (KW)	Interpretation of the KS test for log-transformed data – "Yes" if the null hypothesis holds, "No" if the null hypothesis is rejected.

Levene Test for Homogeneous Variance – Test of equal variance between log-transformed inflow and outflow	
DW stat (Levene)	Reported Durbin Watson statistic for the Levene test of homogeneous variance
Levene SL	Significance level used for the Levene test for homogeneous variance
F ratio (Levene)	Ratio of variances for influent and effluent data
p value (Levene)	Reported p-value for the Levene test.
Indicate Inflow/Outflow Equal Variance (Levene)	Interpretation of the Levene test for homogeneous variance – "Yes" if the null hypothesis (the variance of the two data sets are equal) cannot be rejected, "No" if the null hypothesis can be rejected.

Independent Sample t-Test for Untransformed Inflow and Outflow	
Raw Mean diff	The difference in mean influent and effluent values for untransformed data.
Raw Sep p (t-test)	P-value for the t-test of the raw (untransformed) data assuming that the two samples come from Normally distributed populations with unequal variances (i.e., a separate estimate is used for the variance of each sample).
Indicate Sep Raw Sig Diff (t-test)	Interpretation of separate variance t-test on untransformed data – "No" if the null hypothesis (means are the same) cannot be rejected "Yes" if the null hypothesis can be rejected.
Raw Pool p (t-test)	P-value for the t-test of the raw (untransformed) data assuming that the two samples come from Normally distributed populations with equal variances (i.e., a pooled estimate for the variance of both samples is used in the test).
Indicate Pool Raw Sig Diff (t-test)	Interpretation of pooled variance t-test on untransformed data. "No" if the null hypothesis that the means are the same cannot be rejected "Yes" if the null hypothesis can be rejected.
Independent Sample t-Test for Log-transformed Inflow and Outflow	
LN Mean diff	The difference in mean influent and effluent values for LN transformed data.
LN Sep p (t-test)	P-value for the t-test of the LN transformed data assuming that the two samples come from Normally distributed populations with unequal variances (i.e., a separate estimate is used for the variance of each sample).
Indicate Sep LN Sig Diff (t-test)	Interpretation of pooled variance t-test on the LN transformed data. "No" if the null hypothesis that the means are the same cannot be rejected "Yes" if the null hypothesis can be rejected.
LN Pool p (t-test)	P-value for the t-test of the LN transformed data assuming that the two samples come from Normally distributed populations with equal variances (i.e., a pooled estimate for the variance of both samples is used in the test).
Indicate Pool LN Sig Diff (t-test)	Interpretation of pooled variance t-test on LN transformed data – "No" if the null hypothesis that the means are the same cannot be rejected "Yes" if the null hypothesis can be rejected.